

The logo for proForecaster, consisting of a green square with a white line graph icon.

proForecaster

Professional forecasting software for Excel



User Guide proForecaster 2011 R1

User Guide proForecaster 2011 R1

Copyright © 2011, pro BS UG (haftungsbeschränkt) & Co. KG. All rights reserved.

The program (which includes both the software and documentation) contains proprietary information; it is provided under a license agreement containing restrictions on use and disclosure and is also protected by copyright, patent, and other intellectual and industrial property laws. Reverse engineering, disassembly, or decompilation of the program is prohibited.

proForecaster references the AForge.NET framework and the Accord.NET framework both licensed under GPL v3.

The information contained in this document is subject to change without notice. If you find any problems in the documentation, please report them to us in writing. This document is not warranted to be error-free.

The program is not intended for use in any nuclear, aviation, mass transit, medical, or other inherently dangerous applications. It shall be the licensee's responsibility to take all appropriate fail-safe, backup, redundancy and other measures to ensure the safe use of such applications if the program is used for such purposes, and we disclaim liability for any damages caused by such use of the program.

Microsoft Excel is a registered trademark of the Microsoft Corporation in the U.S. and other countries. Other names may be trademarks of their respective owners.

The program may provide links to Web sites and access to content, products, and services from third parties. pro BS is not responsible for the availability of, or any content provided on, third-party Web sites. You bear all risks associated with the use of such content. If you choose to purchase any products or services from a third party, the relationship is directly between you and the third party. pro BS is not responsible for: (a) the quality of third-party products or services; or (b) fulfilling any of the terms of the agreement with the third party, including delivery of products or services and warranty obligations related to purchased products or services. pro BS is not responsible for any loss or damage of any sort that you may incur from dealing with any third party.

TABLE OF CONTENTS

Welcome	1
Installing proForecaster.....	1
Software Requirements	1
License issues	2
Activate a license.....	2
Forecasting Basics.....	4
Time series forecasting.....	4
Why is forecasting important?	4
proForecaster Professional.....	4
Quick start.....	7
Main menu	7
Forecast.....	7
Regression	7
Welcome	7
Help	7
Info	7
Forecasting time series data	8
Select data.....	8
Forecasting options.....	9
Forecasting result.....	12
Report options	13

Forecasting Multivariate Data	14
Select data.....	14
Select variables.....	14
Regression result	16
Report Options	17
<i>Understanding Time series Forecasting.....</i>	18
A good fitting model.....	18
Time series accuracy measures.....	18
RMSE	18
MAD	18
MAPE.....	19
MPE	19
Theil's U.....	19
LBQ.....	19
Expert rank	20
Smoothing models	20
Forecasting model parameters	21
Moving average.....	22
Exponential smoothing	22
Double moving average	22
Linear smoothing (holt's method).....	22
Damped linear smoothing.....	22
Triple exponential smoothing	22
Additive seasonal method.....	22
Multiplicative seasonal method	23
Additive seasonal trend method.....	23

Multiplicative seasonal trend method	23
Damped additive smoothing	23
Damped multiplicative smoothing	23
Linear Growth	23
Quadratic growth	23
Polynominal growth	23
Neural networks	24
From biology to forecasting	24
Neural network design	25
Number of Input Neurons	25
Number of Neurons in the Hidden layer	25
Number of Neurons in the Output Layer	26
Activation Function	26
Training a neural network	26
Automatic training	26
Manual training	27
Learning Rate	27
Momentum	27
Overfitting a neural network.....	28
PQ Threshold	29
Strip	29
Hybrid forecasting.....	29
<i>Understanding regression</i>	30
Linear regression.....	30
How to assess a good regression model	30
r2	30

Adjusted r2	30
Model significance	30
Variable significance.....	31
Regressing time series data	31
Chart views.....	33
Time series forecasting.....	33
Show Forecast Plot.....	33
Show Residual Plot	34
Show Residual Correlation Plot	34
Show Forecast Value	35
Regression	36
Show Residual Plot	36
Show Forecast Plot.....	37
Show Residual Histogram.....	37
Show Residual Autocorrelation	38
Show Table	39
FAQ	40
How to change the series?	40
How to change the forecast model?	40
How to change the chart view?	40
How to hide a series in the forecast plot?	40
How to zoom into the chart	41
How to adjust forecasts?.....	41
References.....	42

PROFORECASTER

WELCOME

Welcome to proForecaster, the forecasting software which adds advanced time series forecasting functions to Microsoft Excel. proForecaster is an Add In for Microsoft Excel and can be easily used inside the Excel environment to produce time series forecasts for all kind of time series data.

proForecaster helps you to create time series predictions fast and without statistical know-how. This user guide will introduce you to time series forecasting and how to use proForecaster Professional to generate accurate and reliable predictions.

Further information about proForecaster can be found on the proForecaster website www.proforecaster.net

INSTALLING PROFORECASTER

You can download proForecaster from www.proforecaster.net. proForecaster is available in two editions.

- 1) proForecaster Free Edition – free for personal and commercial use
- 2) proForecaster Professional Edition – full features and technical support

When you download the proForecaster Free Edition you automatically receive a 30 days trial license to test the Professional Edition. At the end of the trial period, proForecaster will be downgraded to the Free Edition, if no Professional license is purchased and activated.

- 1) Download proForecaster from www.pro-bs.net
- 2) Open the setup and proForecaster will guide you through the installation process
- 3) After successful installation, proForecaster is available at each Microsoft Excel start-up in the Add In tab.

SOFTWARE REQUIREMENTS

proForecaster requires Microsoft Excel version 2007 or 2010 and the Microsoft .NET Framework version 3.5 to be installed.

LICENSE ISSUES

You can purchase a license from the proForecaster website.
(www.proforecaster.net)

Two professional licensing options are available.

1) Single User License

The software can be used on a single computer for personal and commercial use.

2) Site License

The Software can be used on up to 100 computers of the company that purchased that license.

ACTIVATE A LICENSE

Click on the *License* button in the *Info Menu* and the *License Info* dialog will be shown. In order to activate a license, your computer needs to have access to the internet. The License Manager shows all relevant information such as the status of the license and for whom the license is activated.



Figure 1: License Info

Click on the *Enter License* button and the *Enter License* dialog will open.

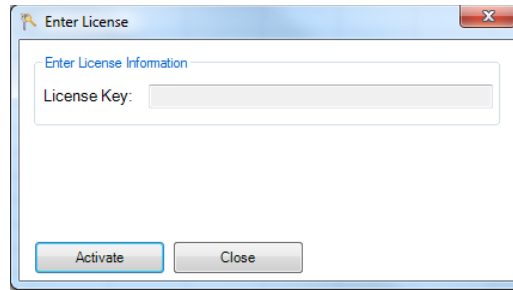


Figure 2: Enter License

In the *License Key* field, enter the license key that was provided to you via email. Click on the *Activate* button and a connection to the pro BS license server will be established. Please note that a Single License can be activated on a single computer as often as needed, in case you installed the software anew.

After successful activation, your license is registered.

FORECASTING BASICS

TIME SERIES FORECASTING

Time series forecasting is an approach to predict future outcomes based upon historical data, whereby different models are applied to the data in order to find the one which best captures trend and seasonal patterns.

Historical data can be anything from quarterly sales recordings to daily stock prices. proForecaster can be used to predict for example:

- Inventory
- Stock prices
- Oil & Gas prices
- Sales demand

The goal is to find a forecasting method that best captures and projects the data into the future.

WHY IS FORECASTING IMPORTANT?

Every organization must try to predict future events. As the timeliness of market actions becomes more important, the need for accurate planning and forecasting is essential to get ahead of competitors. The difference between good and bad forecasting can affect the success of an entire organization.

PROFORECASTER PROFESSIONAL

proForecaster can give you the leading edge in forecasting by providing you with advanced prediction technology that incorporates state-of-the-art developments from artificial intelligence and statistics. proForecaster is graphic oriented and wizard driven, making proForecaster the forecasting tool of choice for business professionals.

Assume you have recorded historical sales data for your main product line for each month over the last 2 years. You want to answer the question “What are the likely sales figures for the next 12 coming months?” Why is this important? Because it allows you to better plan your supplies and minimize your inventory. Generating accurate sales forecasts will help you to save substantial money and streamline your supply chain.

Most historical or time-based data contains an underlying trend or seasonal pattern. However, most historical data also contains random fluctuations (“noise”) that make it difficult to detect these data trends and patterns. proForecaster uses state-of-the-art time series methods to analyze the data and projects them into the future. proForecaster uses two forecasting approaches:

1) **Time series forecasting**

The time series is analysed and projected into the future based on its time series structure.

2) **Regression**

Uses independent variables to forecast a target variable.

Time series forecasting indirectly assumes that all information about the time series is already present in the historical observations, whereas in regression analysis, one or more variables show an effect on the target variable that is to be predicted.

When you run the *Forecast* command on the main proForecaster menu, the time series forecast wizard will be called. The wizard will guide you through the forecasting process in four simple steps.

- 1) You **select** where the **historical data** is located on your sheet.
- 2) You **select** which **forecasting methods** shall be used and how many periods should be forecasted
- 3) You **compare** the different **forecasting models** and judge how they perform on the historical observations
- 4) You **select** where the **predictions** will be inserted into the sheet

To generate predictions, proForecaster has two basic operating modes:

1) **Automatic mode**

proForecaster takes control of the forecasting process and selects a range of candidate models, including *Smoothing Models*, *Growth Functions* and *Neural Network Design*, to find the **best** forecasting model for the data. In the automatic mode you do not have to worry about the technical details of selecting a specific forecasting model and choosing its parameters. proForecaster will do that for you.

2) **Manual mode**

You can select which forecasting models should be applied to the data and which model parameters proForecaster should use.

proForecaster Professional offers three forecasting approaches to forecast time series data.

- 1) **Smoothing models** and **growth function**
- 2) **Artificial neural networks**
- 3) A **hybrid method** where a smoothing model and a neural network are combined to produce forecasts

After the range of models that should be applied to the time series has been selected, four different ranking methods are available to rank each forecasting model. proForecaster provides three commonly used error measures and an Expert ranking method.

- Expert Ranking
- Root Mean Square Error (RMSE)
- Mean Absolute Deviation (MAD)
- Mean Absolute Percent Error (MAPE)

All forecasting models that were tested on the data are ranked according to the ranking method selected and the best model which will yield the most accurate forecast is placed at position one in the ranking. The final forecast shows the most likely continuation of the data. But keep in mind that all forecasting models, no matter how sophisticated they are, depend on the assumption that the pattern found in the data will continue into the future. That means that history will repeat itself to a certain degree.

The further you forecast into the future, the greater is the likelihood that events will diverge from past behavior, and the less confident you can be of the predictions. proForecaster provides confidence intervals for forecasts, that help to gauge how reliable predictions are . With a 90% confidence interval you can be sure that with a certainty of 90% the prediction will lie between the bounds indicated by proForecaster.

QUICK START

MAIN MENU

The proForecaster menu is located in the *Add-Ins* tab in the Microsoft Excel ribbon.

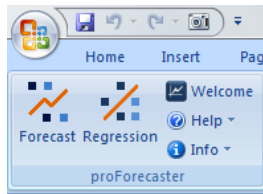


Figure 3: Main Menu

FORECAST

Opens the *Forecast* wizard. This wizard helps you to forecast time series data.

REGRESSION

Opens the *Regression* wizard. This wizard helps you to create forecasts based on one or more variables that influence the variable that shall be predicted.

WELCOME

Opens the *Welcome* screen. The *Welcome* screen provides introductory information and links to resources that may help you to work with proForecaster.

HELP

Opens the *Help* menu where you can find the User Guide, Tutorials and Support, in case you need our help.

INFO

Opens the *Info* menu where you can find information about the proForecaster version and the status of your license.

FORECASTING TIME SERIES DATA

Click in the main menu on the *Forecast* command and the *Forecast* wizard to predict time series data will be shown. The wizard will guide you through the forecasting process in four steps.

SELECT DATA

First you need to select where your data is located. Click into the range field and select the range on the excel sheet or click on the button highlighted in dark red and another window will open, asking you to *Select the range containing the data*. Click into the Excel sheet and select the range of your data.

In the example the sales data in F2:F26 was selected where F2 contains the header for the time series.

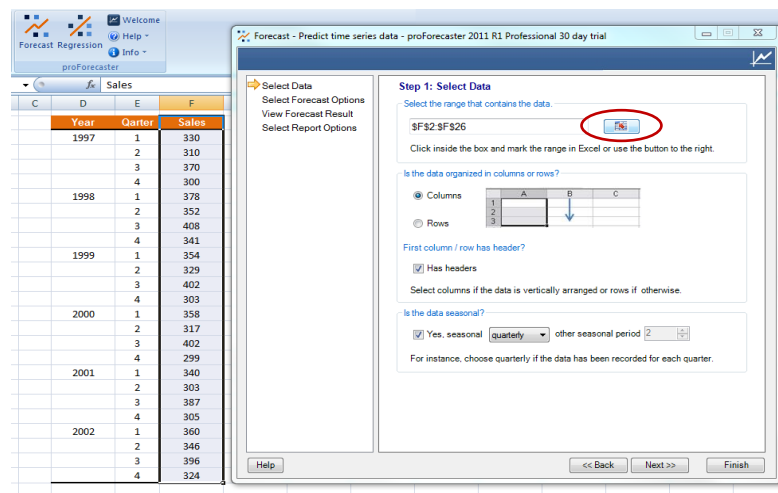


Figure 4: Select Data

After you selected the range containing the time series data, select whether the data is organized in columns or rows. In our case we have just one time series which is arranged in a single column. Note that proForecaster automatically detected that the first row of data contains a header. Select or deselect this option if necessary.

Tip

If you have more than one time series, arrange your data either in rows or columns and select the *Columns* or *Rows* option.

Do not use discontinuous data, that means blank rows or columns. proForecaster will give an error message if a blank cell is encountered in the data range.

Next, proForecaster asks you about the properties of the data, in particular, if the data exhibits **seasonality**. In our case, the sales data was recorded for each month, so check *Yes, seasonal* and choose the kind of seasonality.

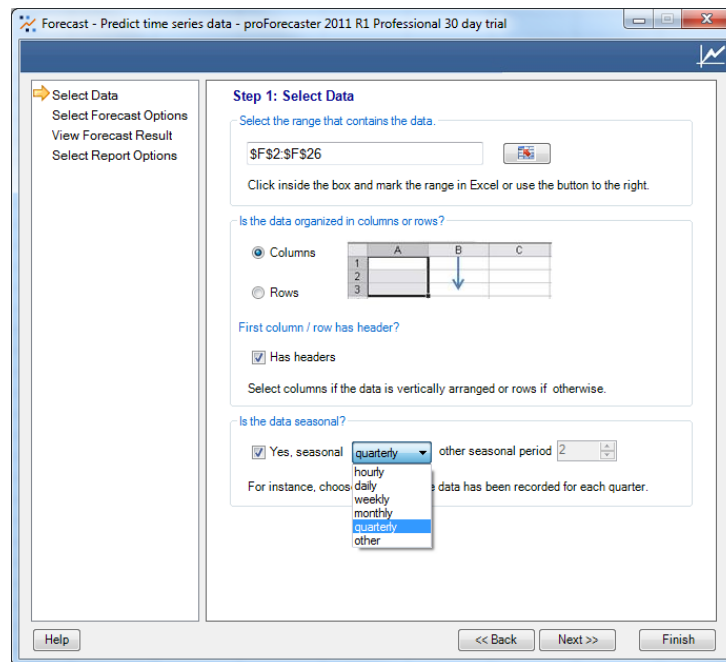


Figure 5: Select Seasonality

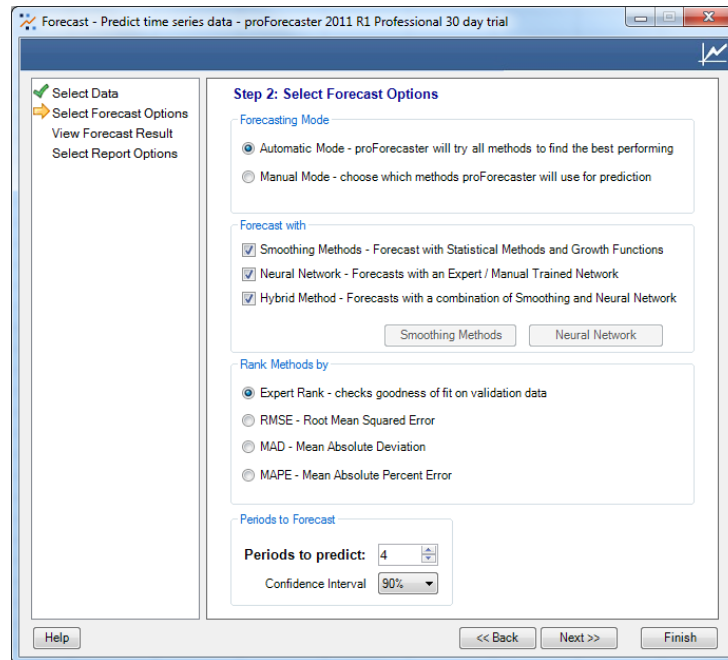
Finally, click on the *Next* button and the next step *Select Forecast Options* will be shown.

FORECASTING OPTIONS

proForecaster can predict data in an **Automatic Mode**, where all technical decisions about which forecasting model to apply and which parameters to select are automatically made by proForecaster through its Expert forecasting engine. The Expert forecasting engine combines expert heuristics and genetic optimization methods to find the right parameters for smoothing models and neural network architectures.

In case you want to select which Smoothing method should be used by proForecaster and you want to design the Neural Network yourself, select the **Manual Mode**.

Figure 6: Forecast Options



proForecaster supports three forecasting approaches

1) Smoothing Methods

These include standard statistical methods such as Linear Smoothing, Moving Average and Growth Functions. proForecaster will apply up to 15 different smoothing methods to your data.

2) Neural Network

Selecting this option, proForecaster will use a Neural Network to forecast your data. In *Automatic Mode*, the network design and training will be performed automatically. proForecaster uses a genetic optimization approach to find the most suitable neural network design for the data.

3) Hybrid Method

Selecting this option, proForecaster will blend the predictions coming from the *Neural Network* and the *Linear Smoothing* method into a new hybrid prediction model.

If you select all three approaches, the likelihood increases that one particular forecasting model will make a *good* fit to your data and will generate accurate forecasts. Please be aware that when you want to forecast several time series in one run and select all three forecasting approaches, a considerable computation time may be required. Especially neural networks are quite computationally intensive.

proForecaster offers four different ranking methods to determine the best forecasting method for the data. *Expert ranking* is the default ranking method for the forecasting models.

1) Expert Ranking

Each model is validated on separate historical data (the time series is divided into 80% of the observations for training the models and 20% for testing the trained models on actual historical data) and a number of statistical properties are examined to test the model's performance.

2) Root Mean Square Error (RMSE)

The RMSE is an absolute error measure that squares the deviations to keep the positive and negative deviations from cancelling out each other. The RMSE is very sensitive to large forecasting errors.

3) Mean Absolute Deviation (MAD)

The MAD averages the distance between each pair of actual and fitted data points.

4) Mean Absolute Percent Error (MAPE)

The MAPE uses absolute values to keep the positive and negative errors from cancelling out each other and uses relative errors to let you compare forecast accuracy between time-series methods.

Select how many periods you want to forecast and whether proForecaster should give a confidence interval for those predictions. In our case we want to forecast 12 months ahead and we want to display confidence bands so that we can be 90% sure. Please note, that the higher the confidence interval, the wider the confidence range for the predictions. proForecaster offers you three commonly used percentages, 90%, 95% and 99%.

Finally, click *Next* and proForecaster will forecast the data. Depending on the number of time series to be forecasted and the number of forecasting models applied to the time series, computation time may range from seconds to several minutes. proForecaster will inform you about the current status of the forecasting run.

FORECASTING RESULT

After the forecasting run, proForecaster displays the forecasting result view. This view displays the forecasting result in graphical form containing the time series, the currently fitted model and the predictions derived from that model, together with method statistics and model parameters.

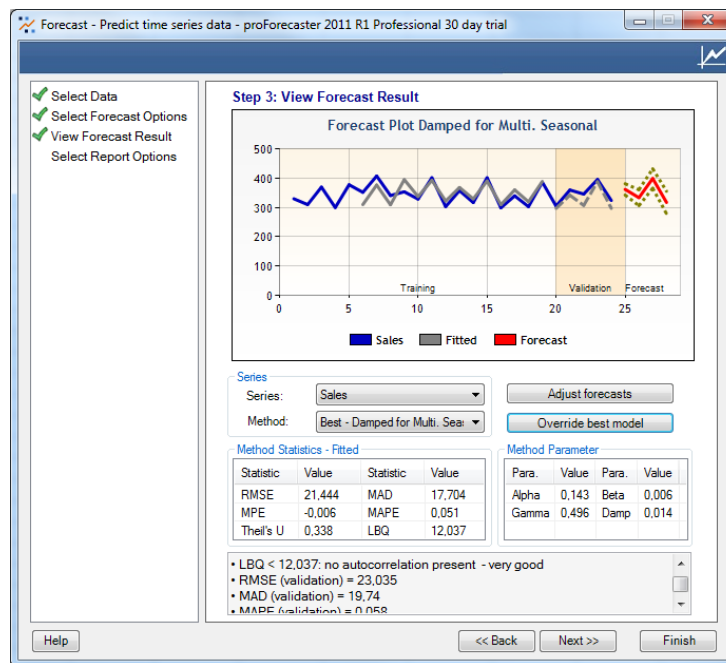


Figure 7: Forecast Result

proForecaster automatically ranks the forecasting models according to the ranking method that was chosen at the previous step.

Each forecasting model that was applied to the data can be selected through the *Method* dropdown. In case you select more than one time series, all time series are listed in the *Series* dropdown.

The *Method Statistics* view displays the main statistics for the selected forecasting model. Note that these statistics are computed based on the fitted forecasts.

The *Method Parameter* view shows the parameters associated with the selected forecasting model. These parameters were either automatically determined by the Expert Forecasting Engine or you entered them in the Manual Mode.

Together with the statistical output, proForecaster gives the main statistics and comments on their values.

The *Adjust forecasts* button opens a new window where forecasts can be adjusted.

The *Override best model* button overrides the best model with the currently selected forecasting model.

REPORT OPTIONS

In the last forecasting step, proForecaster asks you where to insert the predicted values. Per default, the forecasts are inserted at the end of the data range that was selected at step 1.

By checking *other* you can specify in the *Select target cell* box where to insert the forecasts, in case you want them in another place on your sheet. Click into the box and select the cell on your sheet.

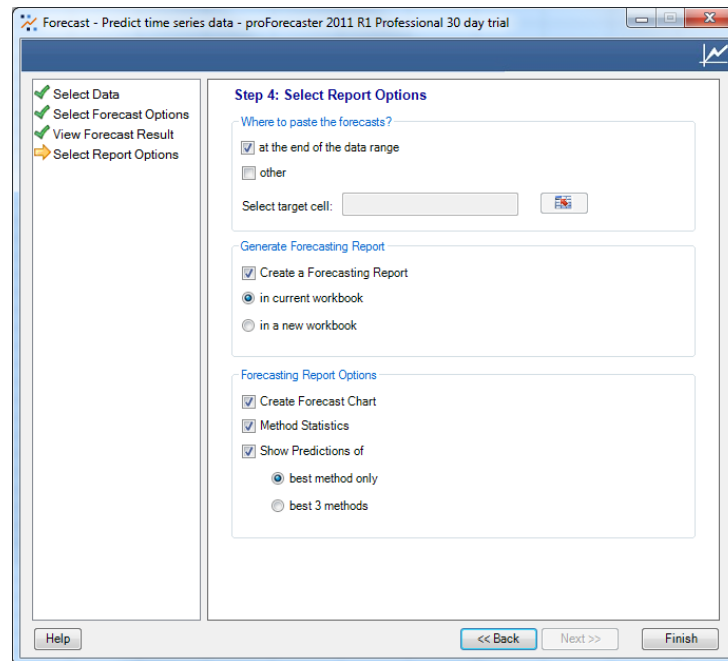


Figure 8: Report Options

proForecaster can create a *Forecasting Report* either in the currently active workbook or in a newly created workbook. You can select which reporting options shall be generated.

- Create Forecast Chart – generates a prediction chart in Excel
- Method Statistics – all method statistics and model parameters are shown
- Show Predictions of – select if you want to display only the best model or 'best' three models

Finally, click on **Finish** and the predictions will be pasted in the Excel sheet together with the Forecasting Report.

FORECASTING MULTIVARIATE DATA

Clicking in the main menu on the *Regression* command will open the Regression wizard for predicting multivariate data. The wizard will guide you through the forecasting process in four steps.

SELECT DATA

Click on the button highlighted in dark red and another window will open, asking you to *Select the range containing the data*. Click into the Excel sheet and select the range of your data.

In the example the sales data in C3:I16 was selected where row 3 contains the header for the data.

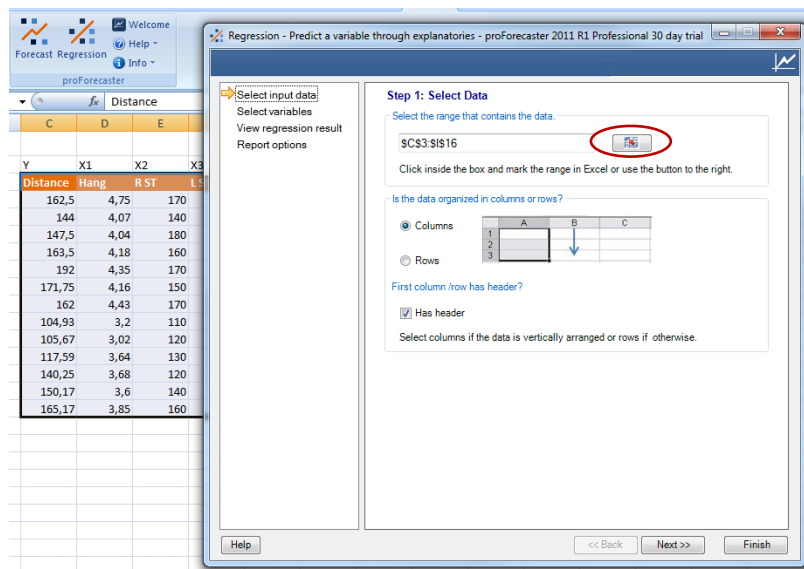


Figure 9: Select Regression Data

Click on the *Next* button and the next step *Select Forecast Options* will be shown.

SELECT VARIABLES

In the *Select Variables* view, you select which variable you want to forecast and which variable or variables should be used to predict the to-be-forecasted variable. Mark the target variables in the *Available variables* list and click on the button  to put them into the respective list.

In order to delete a variable either from the *Variable to be predicted* or the *Explanatory variables* list, use the  button.

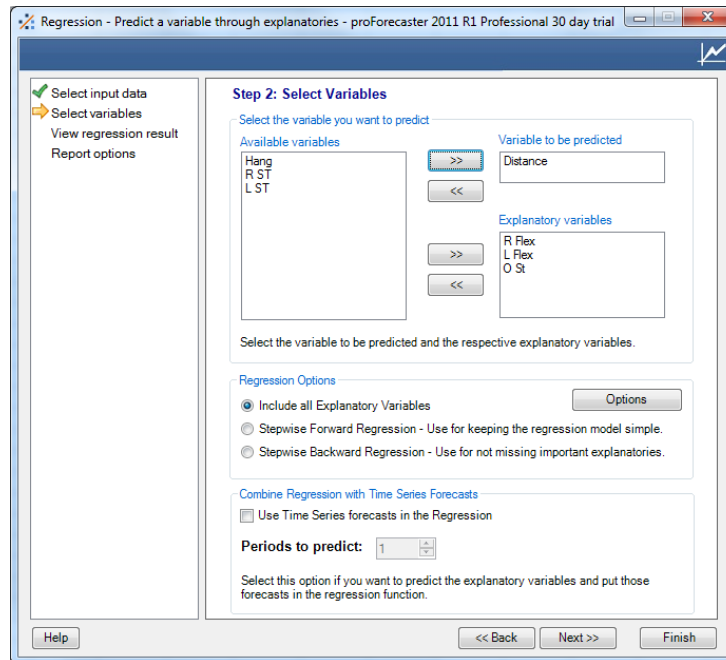


Figure 10: Select Variables

proForecaster supports three Regression options.

1) Include all Explanatory Variables

This is the standard regression method where all variables contained in the *Explanatory variables* list are used to construct the final regression model.

2) Stepwise Forward Regression

This regression method starts with a model containing no variable at all and adds one variable after another to the model. At each run the variable which does the best job of estimating the target variable is included in the model. Use this regression method if you have a lot of potential explanatory variables and want to keep the final regression model as simple as possible.

3) Stepwise Backward Regression

This regression method starts with a full regression model containing all explanatory variables that are listed in the *Explanatory variables* list and removes them one after another until a final model is constructed that includes all important explanatory variables.

In case the *variable to be predicted* is a time series variable, which means it has been recorded over equally spaced time periods, you can select the *Use Time Series forecasts in the Regression* option. This option will automatically generate time series forecasts for the explanatory variables. These forecasts will be used in the final regression model to predict the target variable.

Click on next and proForecaster will construct the regression model.

REGRESSION RESULT

The *Regression Result* view displays the regression result. The *Method Statistics* list all important statistics such as R^2 and the Standard Error of the final regression model. The *Variable Statistics* view shows all explanatory variables and their statistical properties.

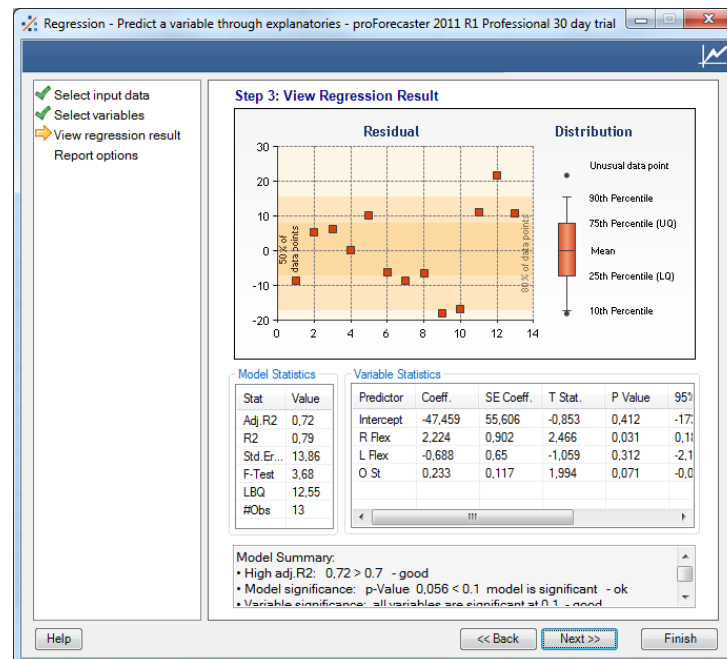


Figure 11: Regression Result

The *Model Summary* summarizes the statistics and interprets the result with respect to the 'goodness of fit' of the regression model.

REPORT OPTIONS

In the *Report Options* view, the report options of the regression result can be selected. You can create a report in the current workbook or create a new workbook.

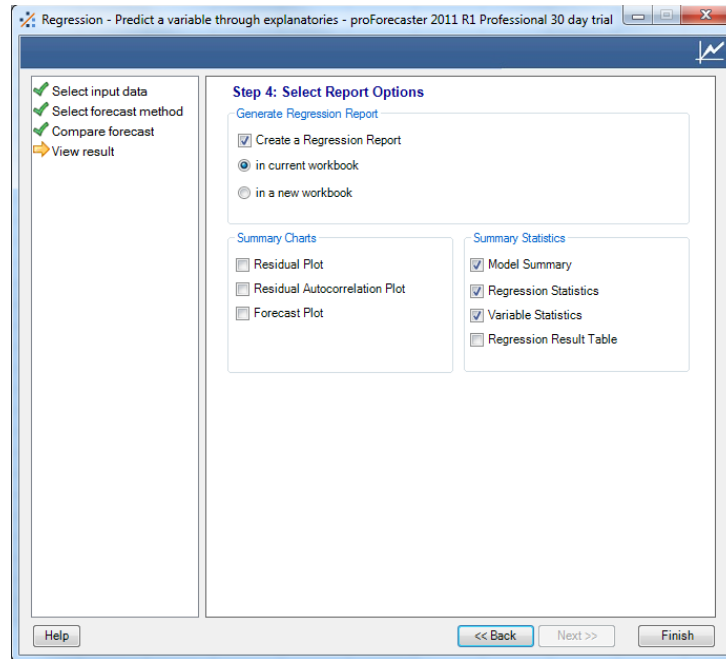


Figure 12: Report Options

Three different charts can be generated.

- Residual Plot – plots the residuals
- Residual Autocorrelation Plot – plots the autocorrelation coefficients for different lags
- Forecast Plot - plots the forecasts for the target variable together with the historical observations

Click *Finish* and the forecasting report will be generated.

UNDERSTANDING TIME SERIES FORECASTING

In this chapter we explore forecasting in more detail and present the main concepts and statistics that help you to judge the goodness of a model.

A GOOD FITTING MODEL

The goal of a time series forecasting process is to find the 'good' forecasting model which is used to generate the predictions. The term 'good' is rather ambiguous. It incorporates several properties. A good forecasting model is one which produces accurate forecasts, measured by some error statistics. Additionally, the forecasting model should capture the structural part of the time series in a way that extrapolations can be made which are robust against random fluctuations. A 'good' forecasting model is consequently a forecasting model which scores high on all those properties and outperforms the competing models.

That said, there is no single forecasting model which performs well on all time series data. The more forecasting models are applied with different statistical properties, the more likely it is, that a model will be considered 'good'.

TIME SERIES ACCURACY MEASURES

A number of accuracy measures have been proposed to summarize the errors/residuals generated by a forecasting model. Most of these measures are based on some function of the difference between the actual value and its predicted value.

RMSE

The Root Mean Squared Error (RMSE) is an often used measure. It sums the squares of the residuals and divides them by the number of observations. Finally, the square root is taken. The RMSE penalizes large forecasting errors due to the squaring of the error.

MAD

The Mean Absolute Deviation (MAD) averages the magnitude of the forecasting absolute values of the errors. The MAD is very useful to measure the forecasting error in the same unit as the time series.

MAPE

The Mean Absolute Percentage Error (MAPE) is calculated by taking the absolute values of the error at each time period and dividing this by the actual observed value. Then the average of these percentage errors is computed. The MAPE indicates how large the forecasting error is compared to the actual values of the time series.

MPE

The Mean Percentage Error (MPE) is calculated similar to the MAPE. The MPE takes the residual at each time period and divides it by the actual value of the series. Finally, the average of the percentage errors is calculated. The MPE helps to decide whether a forecasting model is biased which means if the forecasting model consistently overstates or understates the time series.

THEIL'S U

Theil's U compares the forecasted values to naïve forecasts. A naïve forecast is simply the last observed value taken as the prediction for the next period. The naïve forecast is the simplest to make and the best guess to make when no information is available. A 'good' forecasting model should outperform naïve forecasts.

How to interpret Theil's U:

Theil's U value	Interpretation
More than 1	The forecasting model is worse than guessing
Equal to 1	The forecasting model is as good as guessing
Less than 1	The forecasting model is better than guessing

LBQ

The Ljung-Box Q statistic indicates whether the residuals of a forecasting model show structural patterns. The objective of a 'good' forecasting model is to 'produce' residuals that are randomly distributed. Technically, the LBQ statistic measures whether a set of autocorrelation coefficients are significantly different from autocorrelations that are all zero. Randomly distributed residuals should not be autocorrelated to each other, therefore the LBQ helps to gauge whether some structural part remains in the time series that was not modeled by the forecasting model.

EXPERT RANK

Expert Ranking requires a sufficient number of historical observations, two full cycles of data are usually sufficient. You need, for instance, 24 data points of monthly data to apply the Expert Rank. proForecaster will then divide the observations into a training set (80% of the data) and a validation set (20% of the data).

- 1) The forecasting model parameters will be optimized on the training set
- 2) Forecasts are generated for the validation period
- 3) Forecasts and observations from the validation set are compared
- 4) Validation set error statistics such as RMSE, MAD, Theil's U will be calculated
- 5) The model with the lowest RMSE on the validation set is considered the best model
- 6) All models are rerun to optimize the parameters on all data including the validation set observations

This approach helps to determine how well a model will perform on real world data and should be the default ranking option.

SMOOTHING MODELS

Smoothing models are the classic forecasting methods and have a proven record over many decades. They are simple to apply and provide a robust forecasting accuracy. Every smoothing model relies on an assumption about the time series. Time series are categorized by their fundamental data patterns. Two common data patterns are:

- Trend behavior
- Seasonality

Based on an extended time series classification framework by Pegel (1969) proForecaster provides for every data pattern at least one forecasting model.

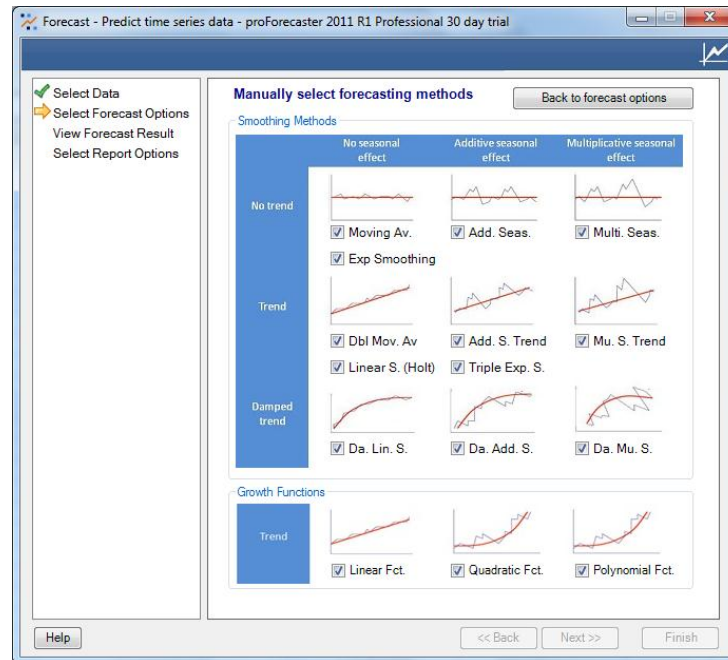


Figure 13: Manual Select Forecasting Methods

Additionally, proForecaster provides standard growth functions such as the linear trend or polynomial growth.

FORECASTING MODEL PARAMETERS

By default, proForecaster optimizes the model parameters through a genetic algorithm where the Mean Squared Error is minimized. This optimization procedure is very robust and provides accurate parameters.

In order to manually select the model parameters, right click on the forecasting model and the *Parameter Options* dialog will open.

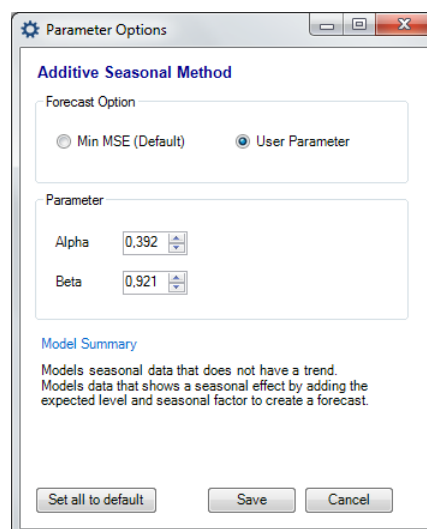


Figure 14: Forecasting Model Parameters

In the *Parameter Options* dialog, you can select how the parameters of the forecasting model should be determined. Use *Min MSE* to let proForecaster determine the optimal parameters. Use *User Parameter* to apply your own parameters to the forecasting model. Click *Save* and the selection will be saved.

The *Set all to default* button applies the default optimization of the parameters to all forecasting models.

MOVING AVERAGE

The Moving Average smoothes out past data by averaging the last two periods and projects that view forward.

EXPONENTIAL SMOOTHING

Exponential smoothing provides an exponentially weighted moving average of all previously observed values. This model is often appropriate for time series with no predictable upward or downward trend.

DOUBLE MOVING AVERAGE

The double moving average computes a first set of moving averages and a second set of moving averages is computed on the first set. This model can model linear trends.

LINEAR SMOOTHING (HOLT'S METHOD)

Holt's exponential smoothing uses a two parameter approach to model data with a trend component. This model is very flexible for trending time series.

DAMPED LINEAR SMOOTHING

Damped linear smoothing is based on linear smoothing and introduces a damping factor into the forecasting model. The damping factor allows to bring conservatism into the trend projections and can therefore model time series which show saturation effects in its trend.

TRIPLE EXPONENTIAL SMOOTHING

Triple exponential smoothing can model quadratic trends and/or seasonality.

ADDITIVE SEASONAL METHOD

The additive seasonal method models data that shows a seasonal effect by adding the expected level and seasonal factor to create a forecast.

MULTIPLICATIVE SEASONAL METHOD

The multiplicative seasonal method models data by multiplying the expected level with the seasonal factor to create a forecast.

ADDITIVE SEASONAL TREND METHOD

This forecasting model is a three parameter model proposed by Holt and Winter to model data with an additive seasonal trend effect.

MULTIPLICATIVE SEASONAL TREND METHOD

This forecasting model is a three parameter model proposed by Holt and Winter to model data with a multiplicative seasonal trend effect.

DAMPED ADDITIVE SMOOTHING

Damped additive smoothing is based on the additive seasonal trend method by Holt and Winter and introduces a damping factor into the forecasting function. This model is useful when the trend should be modeled with a certain degree of conservatism.

DAMPED MULTIPLICATIVE SMOOTHING

Damped multiplicative smoothing is based on the multiplicative seasonal trend method by Holt and Winter and introduces a damping factor into the forecasting function. This model is useful when the trend should be modeled with a certain degree of conservatism.

LINEAR GROWTH

Linear growth is a simple linear regression against time where the sum of the squared residuals is minimized.

QUADRATIC GROWTH

Quadratic growth applies a quadratic regression function to the data.

POLYNOMIAL GROWTH

Polynomial growth is the general growth function. proForecaster supports the application of a polynomial function of a degree up to ten.

NEURAL NETWORKS

Neural networks (NNs) are a powerful and flexible forecasting paradigm. NNs can be used for a variety of forecasting problems. proForecaster uses NNs to model time series data to generate forecasts. This section introduces NNs and how proForecaster uses NNs to create predictions.

FROM BIOLOGY TO FORECASTING

NNs mimic the data processing functionality of the human brain. The human brain consists of billions of neurons that are interconnected. Each neuron is used to store a tiny amount of information and by interconnection, large and complex information can be stored and processed.

For time series forecasting, an artificial neural network is created which takes a number of historical observations as input to predict a value one step ahead. Several thousand runs are performed to update the weights with which each input influences the one step ahead prediction. The objective is to minimize the overall forecasting error.

NNs possess an interesting property which makes them especially suitable for time series forecasting. NNs do not assume a certain underlying data generation process of the time series. Unlike smoothing models, NNs are not bound to data patterns such as trend or seasonality. NNs learn the pattern directly from the data and can model even non-linear data patterns. The interdependencies in our complex world are seldom linear, in fact non-linearity is found in many business forecasting problems. A NN can approximate any linear and non-linear continuous function to any level of accuracy, which makes them a must have in forecasting software.

proForecaster uses a three layer neural network consisting of an input, a hidden layer and an output layer. The NN is trained through backpropagation, where at each training epoch the information coming from the input neurons is multiplied by its assigned weight. The result is fed into the activation function which fires when a certain threshold is reached and then sends the signal to the output neurons. At each run, the forecasting error is calculated and fed back into the NN.

NEURAL NETWORK DESIGN

The challenge of applying a NN to a time series forecasting problem is to choose the 'right' design of the NN. This requires to determine how many neurons should be in the input, hidden and output layer. Furthermore, if a neuron is presented with a signal coming from other neurons, the activation function tells the neuron how to react to that signal. proForecaster provides four commonly used activation functions to choose from.

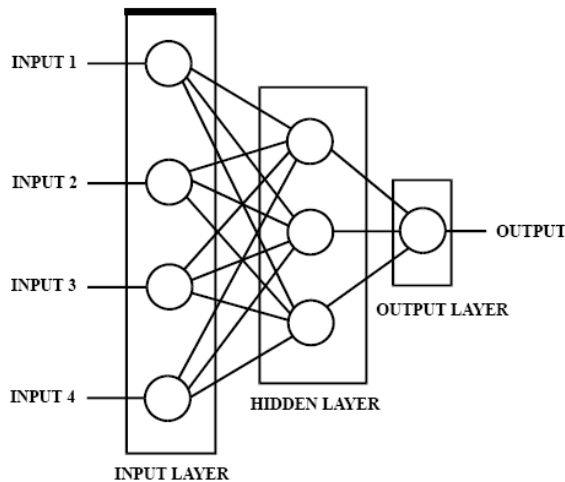


Figure 15: Network Architecture

NUMBER OF INPUT NEURONS

This is the first choice to make; how many past periods should be used to predict a future value. If you select the number of input neurons to be two then the forecasted value is a function of its past two predecessors.

NUMBER OF NEURONS IN THE HIDDEN LAYER

In the hidden layer, the data processing takes place. Selecting the number of neurons in the hidden layer is a non-trivial task. The more neurons are in the hidden layer the more complex data patterns can be learned by the NN. But that does not mean that the NN gets smarter by choosing a high number of neurons in the hidden layer, in fact the higher the number of neurons in the hidden layer the higher the risk of overfitting. Some studies suggest that the amount of neurons in the hidden layer is a function of the number of neurons in the input layer.

NUMBER OF NEURONS IN THE OUTPUT LAYER

This is the easiest selection, since we use the NN to make a one step ahead forecast which means the output layer has a single neuron. proForecaster always trains the NN to make a one step ahead forecast through a sliding window approach.

ACTIVATION FUNCTION

proForecaster offers four activation functions:

- Bipolar Sigmoid
- Sigmoid
- Hyperbolic Tangent
- Semi-linear

TRAINING A NEURAL NETWORK

Before a NN can be used to forecast a time series, it has to be trained on its historical observations. For each NN you want to forecast you need one trained NN. The general idea of training a NN is to present it with data from the time series such that it can learn the data pattern in a way that it can generate predictions.

Please note that each time a NN is trained on one time series with identical parameters, the NN may lead to different forecasting results. That is due to the random initialization of the neurons each time the neural network is trained. Therefore, even with identical parameters, two NNs trained on the same time series may lead to slightly different forecasts.

AUTOMATIC TRAINING

Through its 'Automatic' forecasting mode, proForecaster provides the possibility to use NNs for predictions without requiring any knowledge about the NN design. proForecaster will select the NN architecture and all other parameters automatically. proForecaster uses expert heuristics and a genetic optimization approach to train several neural networks with different parameters. Finally, one 'surviving' NN is chosen for the predictions. This approach allows to evaluate several different NN architectures and find the most suitable one for the time series to be forecasted.

MANUAL TRAINING

proForecaster supports the manual training of NNs. Select the 'Manual' forecasting mode and the button *Neural Network* will be enabled. Click on that button and the *Manual train Neural Network* view will be shown.

Figure 16: Manual Training

Forecast - Predict time series data - proForecaster 2011 R1 Professional 30 day trial

Manually train Neural Network [Back to forecast options](#)

Train Neural Network for Series

Series:
Select the series for which you want to train the neural network.

Network Architecture

of Input Layer Neurons Learning rate
of Hidden Layer Neurons Momentum
Activation Function #Trials
Choose how the neural network should be designed.

Stop Training Condition

☒ Cross validate PQ Loss Threshold Strip
Learning Error General Loss
Validation Error PQ

Training Error

Epoch
MSE
Error
[Save NN for prediction](#)
[Start training](#)

Help << Back Next >> Finish

In the *Manual train Neural Network* view you can select for which time series you want to manually train a NN.

Select the NN architecture, i.e. the number of neurons in the input and hidden layer. (see Neural Network Design) Furthermore, select for how many epochs you want to train the NN. If no stopping condition is applied to NN training, proForecaster will train the NN until the last epoch finished.

LEARNING RATE

The learning rate of a NN controls how quickly the weight of the neuron changes. A too high learning rate may lead to large oscillations of the NN. Whereas a too small learning rate may lead the NN to be stuck in a local minimum and not finding the lowest forecasting error possible.

MOMENTUM

The momentum controls the tendency of a weight to change the direction. That means, that each weight remembers the weight where it comes from and if it increased or decreased, and the momentum tries to keep the weight in that direction. With a small momentum, weights are allowed to change more freely,

whereas a high momentum forces the weight more into a certain direction. Thus a NN responds more slowly to new training data.

OVERFITTING A NEURAL NETWORK

As already mentioned, a NN can approximate any function to any degree of accuracy. This leads to the problem of overfitting, where a NN captures the time series perfectly but does not have enough generalization capacity to make accurate predictions. Each time series consists of some structural element and some degree of noise. A 'good' forecasting model should only approximate the structural part and not the noise. Overfitted NNs can be avoided by applying a stopping condition to the NN training. This condition is used to detect overfitting during the training process. When a NN begins to overfit the data, the training is automatically stopped.

Figure 18 shows the evolution of the training error and the validation error over time. The idea is to split the time series into a training set used to train the neural network and a validation set to validate its predictions. At the start of the training, both, the training and the validation error decrease rapidly. After some epochs, the NN overfits, i.e. the training error still decreases but the validation error changes direction and rises.

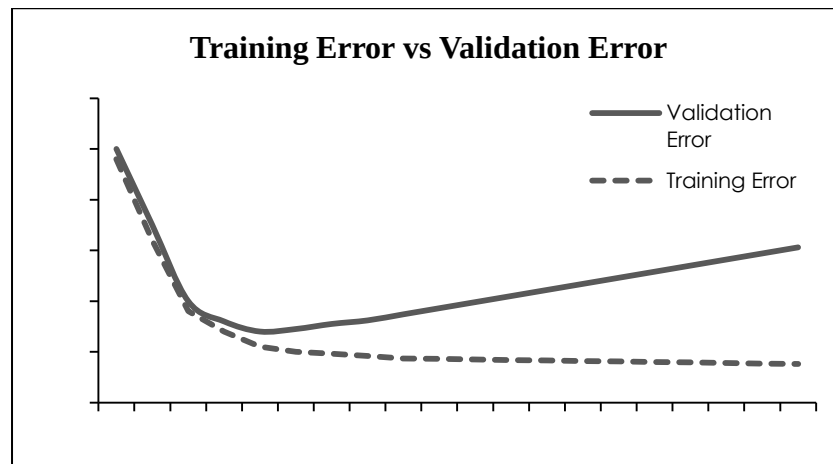


Figure 17: Training vs. Validation Error

To overcome overfitting, a formal stopping condition for NN training is required. proForecaster implements the advanced stopping through applying a *PQ Threshold* and a *Strip* parameter to the training process. The concept is that as soon as the generalization loss exceeds a certain threshold, the training is stopped. Generalization Loss is the current error of the validation set divided by the lowest validation error obtained so far.

However, we want the training progress to continue as long as the training error decreases rapidly. Overfitting occurs when the training error decreases only slowly. For that reason, we define the quotient of generalization loss and progress (PQ) as the generalization loss divided by the training progress made during k epochs. The progress allows to measure whether the training error decreases slowly and whether the training is approaching overfitting.

PQ THRESHOLD

The threshold when the quotient of generalization loss and progress is used to indicate overfitting. In general, a PQ Threshold of 1.5 is often used.

STRIP

The number of training epochs that are used to measure the training progress. A strip of 5 epochs is recommended.

HYBRID FORECASTING

Hybrid forecasting is an approach to combine different forecasting approaches and techniques into a single forecasting method. The idea behind hybrid forecasting is that different complementary properties of single forecasting methods are combined to create forecasts that enjoy a high robustness.

proForecaster blends the forecasts derived from Linear Smoothing and a Neural Network into the hybrid model. The weight of each model is determined by finding the smallest absolute error of the combined forecasts.

$$\text{hybrid forecast} = \alpha \text{ linear forecast} + (1-\alpha)\text{NN forecast}$$

Hybrid forecasting can be especially interesting for financial time series forecasts. These time series often show linear and non-linear data patterns.

UNDERSTANDING REGRESSION

Regression is suitable for forecasting applications when there is a relationship between the to-be-forecasted variable and other independent variables. In regression a mathematical relationship is built to describe how strong the influence of the independent variables is on the target variable.

LINEAR REGRESSION

proForecaster supports multivariate forecasting through linear regression. Here a linear relationship is built between predictor variables and the to-be-forecasted variable. For a neat introduction to regression we recommend the book *Intermediate Statistics for Dummies* by Deborah Rumsey.

HOW TO ASSESS A GOOD REGRESSION MODEL

In order to judge whether a regression model is suited to generate forecasts, its statistical properties and the distribution of the residuals have to be examined.

R²

Measures how well the variability of the target variable is explained by the predictor variables. R^2 can range from 0 and 1. Where 1 is a perfect relationship and 0 is no relationship at all. An R^2 of 0.7 and higher is considered to be a quite good value.

ADJUSTED R²

One interesting feature about R^2 is that the more predictor variables we add to the regression model, the higher the R^2 becomes. This is due to the fact the more variables we add, the more information can be used to predict the target variable. However, this better explanation may be only due to pure chance.

Adjusted R^2 accounts for this and only increases if the new variable added is providing information that would not be expected by chance.

MODEL SIGNIFICANCE

If the regression model is significant can be determined by the F Test that is shown in the *Model Statistics* view. proForecaster automatically interprets the F Test statistic and comments whether the whole model is statistically significant or not.

VARIABLE SIGNIFICANCE

proForecaster shows whether a variable is statistically significant by displaying the p-Value. If the p-Value is lower than 0.05 than the variable is significant with a confidence of 95%. proForecaster uses the alpha value specified in the Regression Options view to determine the significance.

REGRESSING TIME SERIES DATA

proForecaster can automatically generate time series predictions for the independent variables and insert them into the regression function. This helps to short cut the forecasting process.

Select *Use time series forecasts in the regression* and enter the number of periods you want to forecast.

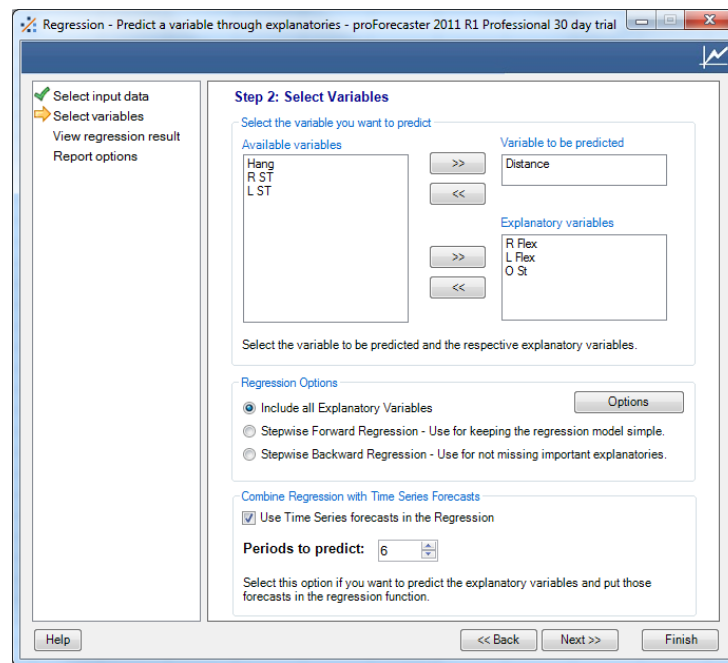


Figure 18: Training vs. Validation Error

In the *Regression Result* view, the forecasting plot will show the historical data of the target variable, the historical fitted data created by the regression and the forecasts.

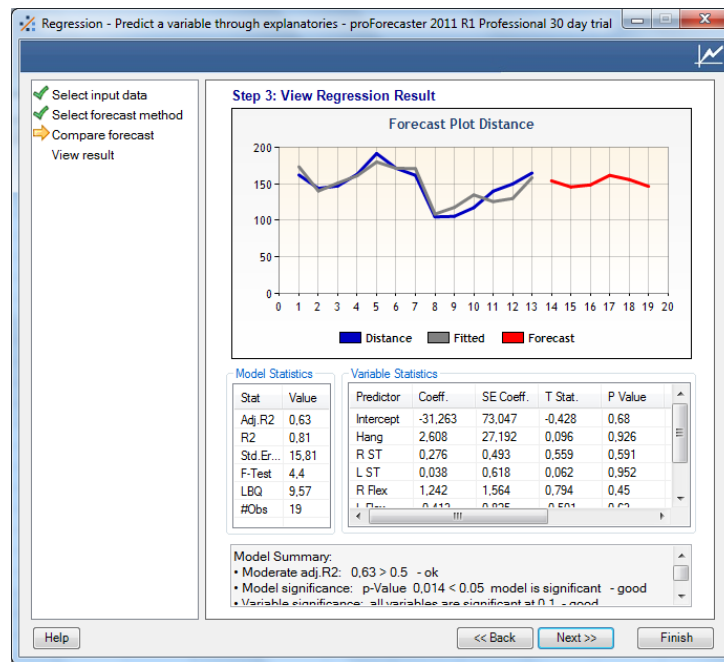


Figure 19: Time Series Regression Result

CHART VIEWS

proForecaster provides the visual inspection of the residuals through different charts. Charts help to check whether a time series forecasting or a regression model is a good fit to the data.

To change the chart view right click into the chart and a menu opens with different charts to select.

TIME SERIES FORECASTING

proForecaster provides four different charts in the *Forecast Result* view.

SHOW FORECAST PLOT

Displays the time series, the fitted values and the forecasts for a forecasting model.

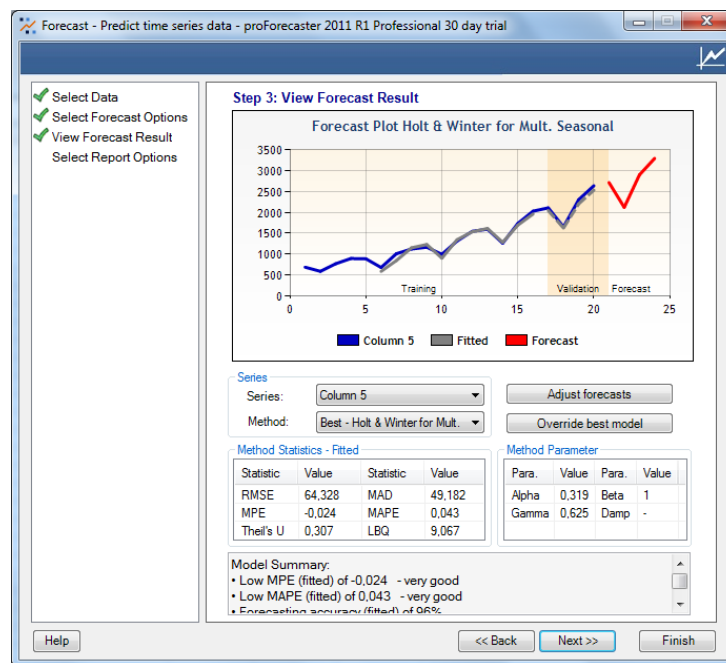


Figure 20: Forecast Plot

SHOW RESIDUAL PLOT

Displays the residuals of the current forecasting model. This chart is useful to determine whether a structural part is present in the residuals. Good forecasting models produce randomly distributed residuals.

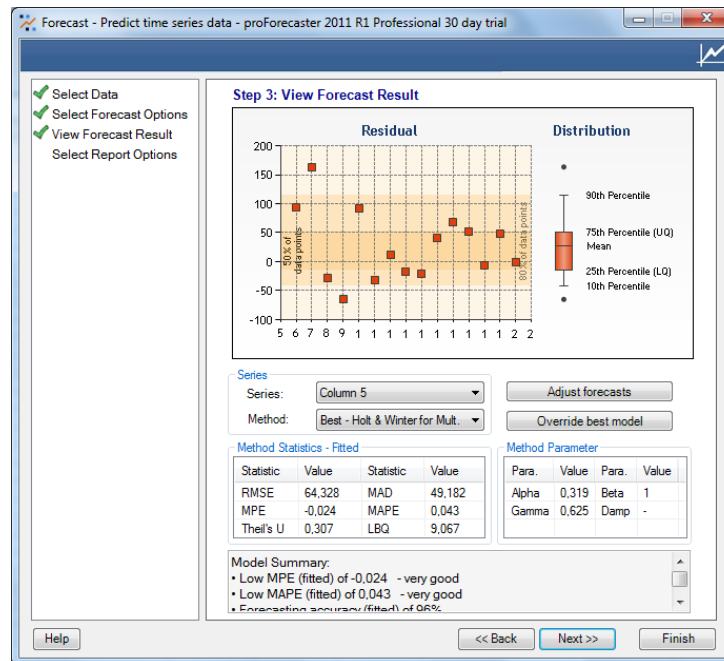
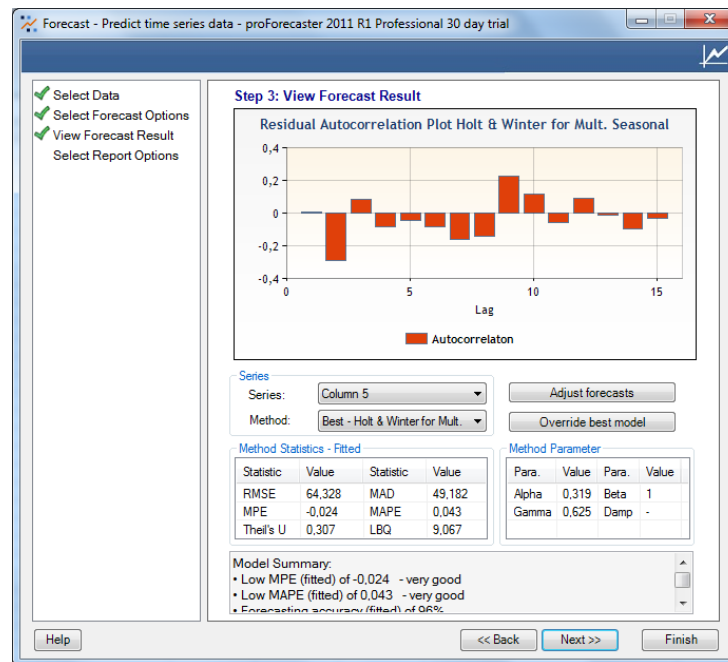


Figure 21: Residual Plot

SHOW RESIDUAL CORRELATION PLOT

Displays the autocorrelation coefficients at different lags of the residuals of the current forecasting model. This chart helps to determine whether the forecasting model is a good fit to the data. High autocorrelation coefficients indicate that a data pattern is present in the time series that is not adequately captured by the forecasting model. The autocorrelation coefficient can be in the range of -1 to +1. An autocorrelation coefficient of ± 0.6 for a lag can be an indication of autocorrelation.

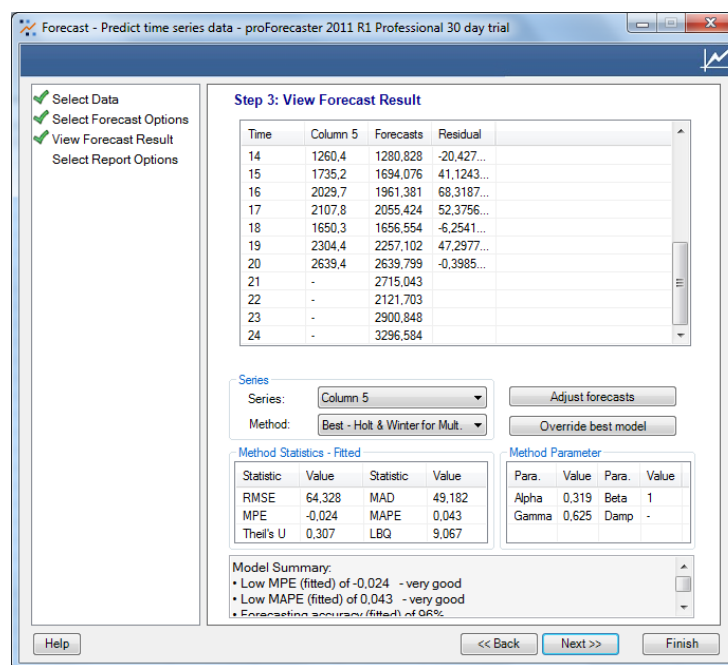
Figure 22: Residual Autocorrelation Plot



SHOW FORECAST VALUE

Displays a data table showing the time series values and the forecast values

Figure 23: Forecast Values View



REGRESSION

proForecaster provides five different charts on the regression result view.

SHOW RESIDUAL PLOT

Displays the residuals of the current regression model. This chart is useful to determine whether a structural part is present in the residuals. An assumption of a linear regression model is that the residuals are randomly distributed.

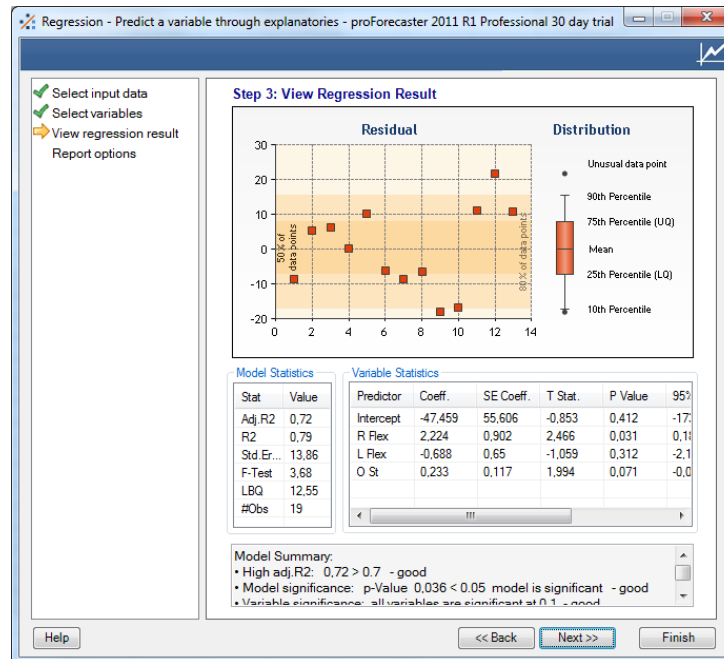


Figure 24: Residual Plot
Regression

SHOW FORECAST PLOT

Displays the target variable, the fitted values created through the regression model and the forecasts. This chart is only available if the *Use time series forecasts in the regression* is chosen.

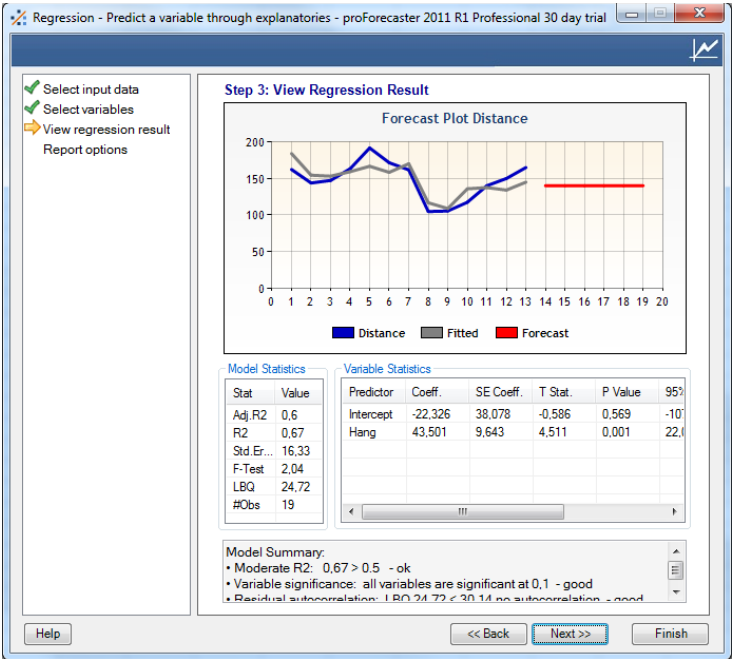
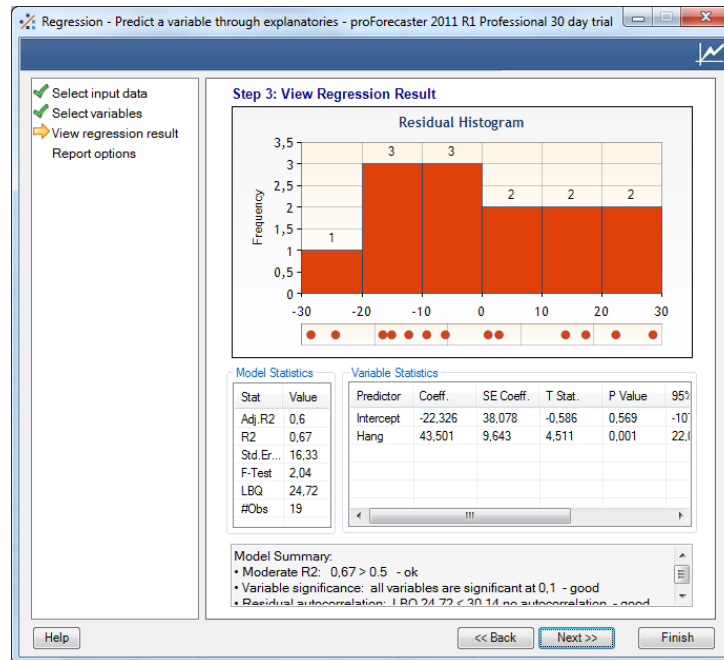


Figure 25: Forecast Plot
Regression

SHOW RESIDUAL HISTOGRAM

The residual histogram shows the distribution of the residuals. This graph helps to identify whether the residuals show a normal distribution as is assumed by the linear regression model.

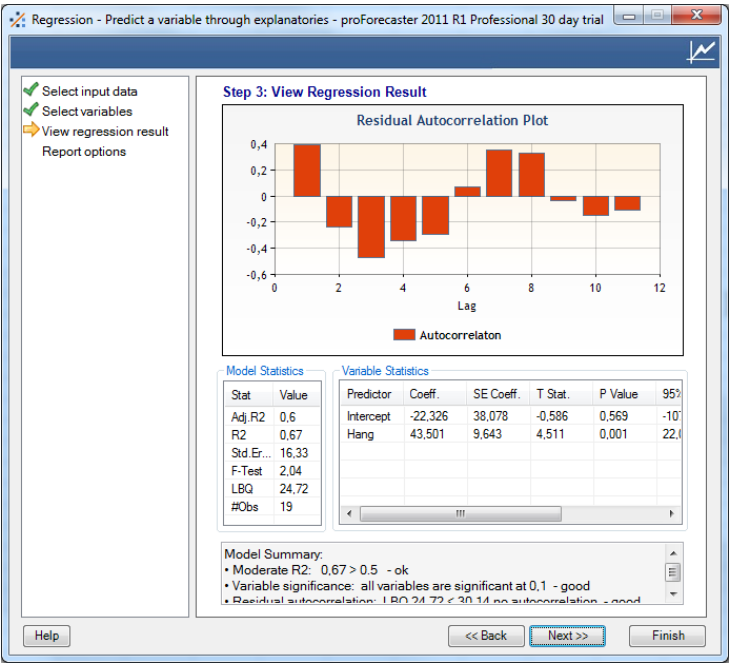
Figure 26: Residual Histogram



SHOW RESIDUAL AUTOCORRELATION

Displays the autocorrelation coefficients at different lags of the residuals generated by the regression model. This chart helps to determine whether autocorrelation is present in the residuals. High autocorrelation coefficients indicate that a data pattern is present in the residuals, thus violating the assumptions of the linear regression model. The autocorrelation coefficient can be in the range of -1 to +1. An autocorrelation coefficient of ± 0.6 for a lag can be an indication of autocorrelation.

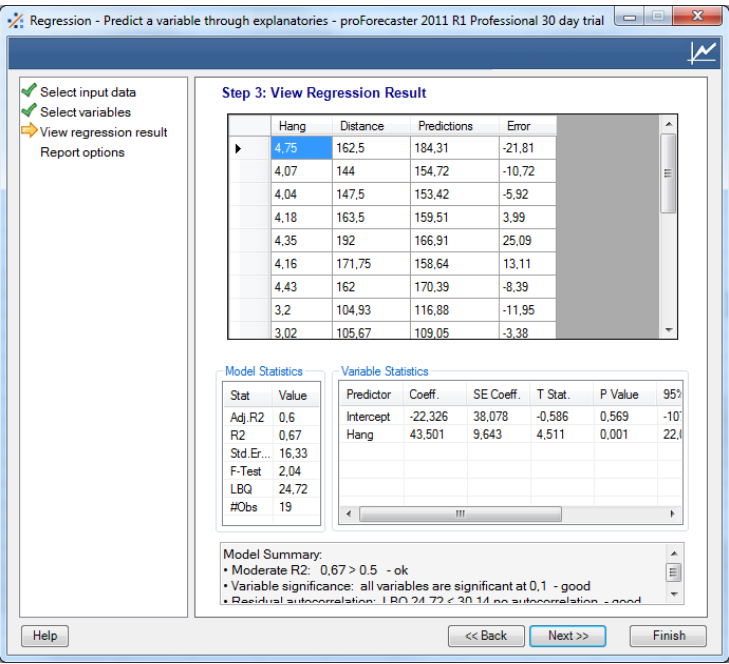
Figure 27: Residual Autocorrelation Regression



SHOW TABLE

Displays a data table showing all independent variables, the target variable and the residuals for the current regression.

Figure 28: Table Regression Values



FAQ

HOW TO CHANGE THE SERIES?

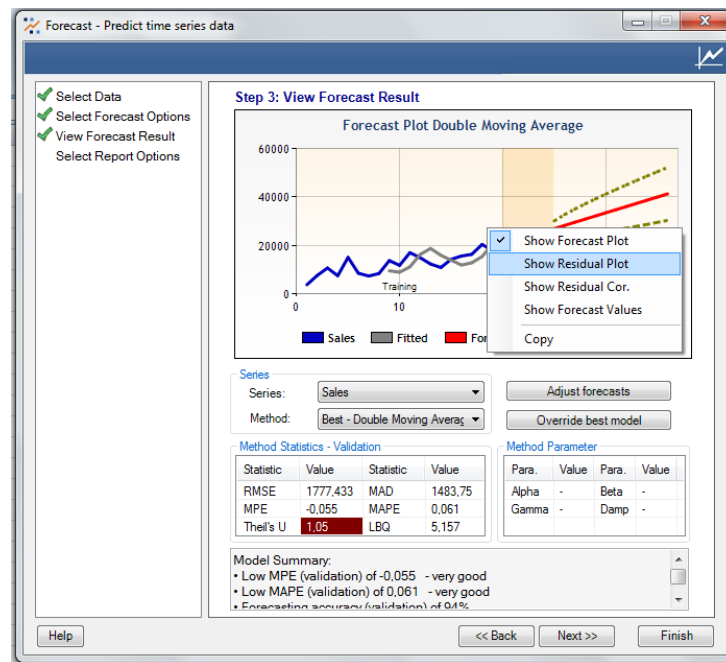
In the *Forecasting Result* view, change the value in the *Series* dropdown.

HOW TO CHANGE THE FORECAST MODEL?

In the *Forecasting Result* view, change the value in the *Method* dropdown.

HOW TO CHANGE THE CHART VIEW?

Right click into the chart area and the context menu will be shown.



HOW TO HIDE A SERIES IN THE FORECAST PLOT?

Click on the label name in the legend of the chart to either display or hide the series.

HOW TO ZOOM INTO THE CHART?

Just click into the chart and draw a rectangle, this will zoom into the area. Use the minus and plus sign to change the scale as needed.

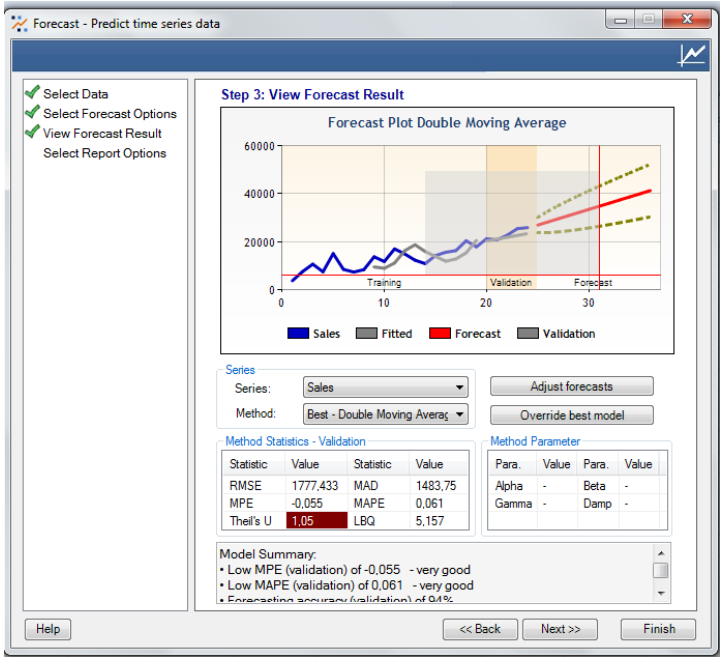


Figure 30: Zoom Into A Chart

HOW TO ADJUST FORECASTS?

Forecasts can be directly adjusted inside the *Adjust Forecast* dialog. You can adjust the forecasts by a certain value. You can round the values and set lower and upper bounds.

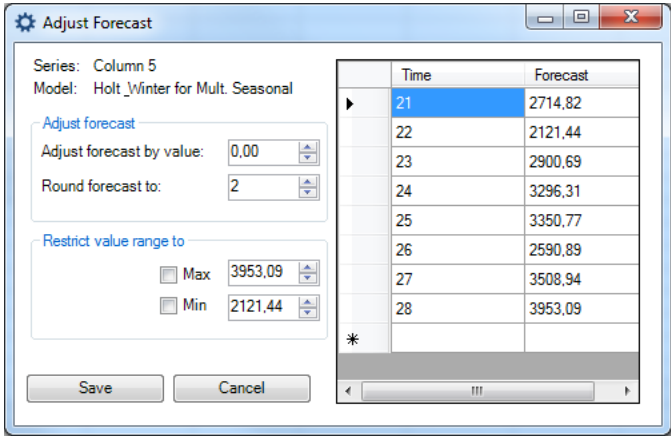


Figure 31: Adjust Forecast

REFERENCES

Bishop, Christopher M. (2006), *Pattern Recognition and Machine Learning*, Springer 2006

Hanke, John E. et al (2001), *Business Forecasting*, 7. Ed Prentice Hall International Inc., 2001

Linoff, Gordon S. and Berry, Michael J.A. (2011), *Data Mining Techniques*, 3. Ed. Wiley Publishing Inc. 2011

Makridakis, Spyros et al. (1994), *Forecasting: Methods and Applications*, 2.Ed John Wiley & Sons, 1994

Pegels, Carl C. (1969), Exponential Forecasting: Some New Variations, *Management Science* Vol.12 No. 5, p.311-315, 1969

Ragsdale, Cliff T. (2004), *Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science*, 4. Ed Thomson South-Western, 2004

Rumsey, Deborah (2007), *Intermediate Statistics for Dummies*, Wiley Publishing Inc. 2007